# Measuring Health and Ageing

Yannick Schindler

EIT Oxford

25 November 2025

# Roadmap for today

1. **Conceptualising health:** Capital (Grossman) vs. deficits (Strulik/Dalgaard)
2. **Self-reported measure of health:** Reporting bias and anchoring vignettes
3. **Objective measurement:** Frailty, chronic disease indices, epigenetic clocks
4. **New opportunities:** Modern ML / LLMs in health data
5. **Welfare & policy:** QALYs, HTA design, and fiscal feedback

# WHAT IS HEALTH?

# Different definitions

- **WHO definition:** “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity”

- **Capabilities perspective (Sen):**
  - Health as the capability to do and be certain things.

- **Economist’s typical definition?**

## Before trying to measure health, let's try to conceptualize it (via economist's lens)

- It seems we need an object in our models $H_t$ that, at a minimum:
  - Evolves over time.
  - Responds to investments and behaviours.
  - Affects utility (directly or indirectly) and survival.

- Two paradigms to get us started:
  1. Grossman: health as a **capital stock**.
  2. Strulik: health as **accumulated deficits / frailty**.

# Grossman (1972): health as capital

- Notion of human capital already well-explored: Becker (1967); Ben-Porath (1967)
  - Formal schooling
  - On-the-job training

- If $H \uparrow$ simply means productivity/wages $\uparrow$, then can use those existing frameworks.

- Grossman: health capital $\neq$ human capital
  - Health capital affects time endowments (time alive, time spent in good health)

## Grossman (1972): health as capital (stylized)

- Health stock $H_t$:

$$\dot{H}_t = I_t - \delta H_t$$

  where $I_t$ is health investment, $\delta$ is depreciation rate.

- Agent gets direct utility from "healthy days":

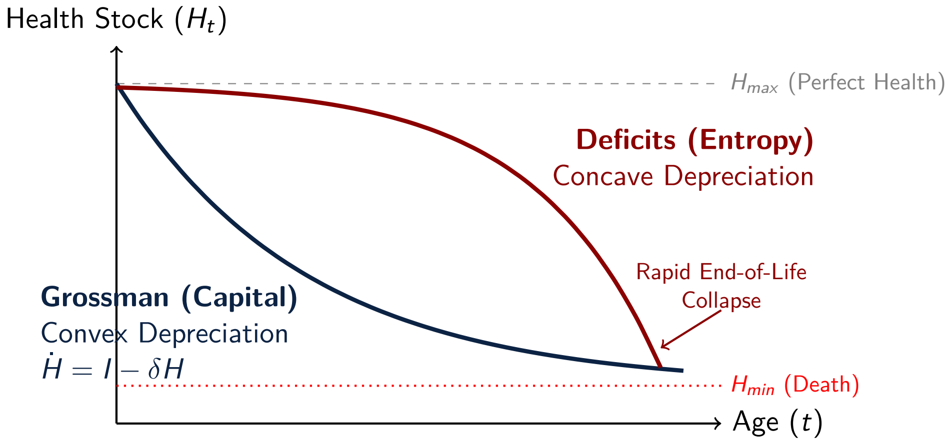
$$U_t = u(C_t, H_t)$$

- Health affects length of life. Death happens when $H_t \leq H_{min}$

- Sick days deplete time endowment:

$$\Omega = T_{work} + T_{invest} + T_{sick}$$

# Economic depreciation vs. biological entropy




Note: The Convex path implies rapid initial loss (depreciation). The Concave path implies slow initial loss followed by accelerating damage (entropy), which better fits biological data.

# Limitations of the health capital view

- Health deteriorates rapidly then slowly (but biology literature suggests the opposite)
- Hard to match empirical mortality patterns (Gompertz law).
  - Can overcome this by making $\delta$ increase with time
  - But biology suggests that "health deterioration rate" depends on health state, not chronological time.
- Health improvements can, in principle, fully offset ageing
  - Large $I_t$ can restore a 100-year-old to same health as 20-year-old
- Motivates moving to **deficit accumulation** frameworks.

- Health represented by a **deficit index** $D_t \in [0, 1]$:

$$\dot{D}_t = \mu D_t + \varepsilon_t - \phi(I_t)$$

  - $\mu D_t$: entropic accumulation of damage.
  - $\varepsilon_t$: shocks (disease, accidents).
  - $\phi(I_t)$: effect of health investment on **slowing damage**

- Mortality hazard increasing in $D_t$:

$$m(t) = m_0 e^{\gamma D_t} \quad \text{(or similar)}$$

- Matches **Gompertz law**: human mortality doubles approx. every 8 years
- **Entropic ageing**: health deficits accumulate faster the more damage we have.

# Ok so how do we try to measure health?

- Recall, we want something that:
  - Evolves over time.
  - Responds to investments and behaviours.
  - Affects utility (directly or indirectly) and survival.

- My take: $H_t$ (or $D_t$) **latent state** of the body and mind.

- We observe noisy **signals**: self-reports, diagnoses, biomarkers, mortality.

- **Challenge:** use the signals to estimate the latent health state

# Some things we observe

- Physical: mobility, chronic disease, organ function.
- Physiological health: biomarkers, inflammation, metabolic status.
- Cognitive health: memory, processing speed, executive function.
- Mental health: depression, anxiety.
- Functional status: (ADLs/IADLs, work capacity).
- Question: can we collapse this into a **single index**? Should we?

# Self-reported health (SRH)

- Typical survey question:

  *"In general, would you say your health is: excellent, very good, good, fair, or poor?"*

- Advantages:
  - Cheap, widely available, strongly predictive of mortality.
- Problems:
  - State dependent reporting
  - Reporting heterogeneity across gender, SES (differential item functioning)
  - Justification bias

# Non-classical measurement error in health

- Ideal setting: observed health $H^{\text{SRH}} = H^* + \eta$ with $\eta$ classical white noise measurement error
  - Coefficint on $H^{obs}$ attenuated

- In practice:

$$H^{\text{SRH}} = H^* + \eta, \quad \eta \text{ correlated with } H^*,\ X,\ \varepsilon$$

- Implications:
  - Bias when regressing labour supply, wages, disability insurance (DI) payments on SRH.
  - Bias direction is not alwyas obvious ex ante.

# Justification bias

- **Issue:** Agents with low labour supply or high DI receipt may **down-report** health to rationalise choices.
- Example:
  - Individual leaves the labour force early.
  - Reports worse health after exit, even if clinical measures unchanged.
- Econometric consequence:
  - Health appears more strongly related to non-employment than it truly is.
  - Reverse causality between labour status and reported health.

- **Issue:** Reporting may vary across genders, SES, etc.
- **Solution:** Respondents are asked to **rate hypothetical individuals** (vignettes) with described health states.
- All respondents read same scenario and then asses health:

  **"John suffers from back pain that makes it uncomfortable to sit for long periods, but he is able to walk around the block"**

- Under assumptions, we can recover group-specific thresholds.
  - Vignette equivalence: respondents interpret the scenario in the same way (e.g. no issues in translation across languages)
  - Response consistency: respondents use same scale for John as they use for themselves.

## Frailty indeces

- **Definition:** A continuous measure of biological ageing based on the proportion of deficits accumulated (Mitnitski & Rockwood, 2001).

- **Construction:** Given a set of $N$ symptoms, signs, or functional impairments (where $d_i = 1$ if present):

$$FI_i = \frac{\sum_{j=1}^{N} d_{i,j}}{N}$$

  - $FI \in [0, 1]$ (0 is perfect health, 1 is total system failure).
  - Vast majority die before going above 70

- **Interesting statistical property:**
  - You need $N \approx 30$ variables.
  - *Crucially:* It does not matter *which* variables you include (e.g., "trouble walking" vs. "trouble feeding self'). See e.g. Searly et al. (2008).
  - Index seems to capture something latent, not a specific pathology.

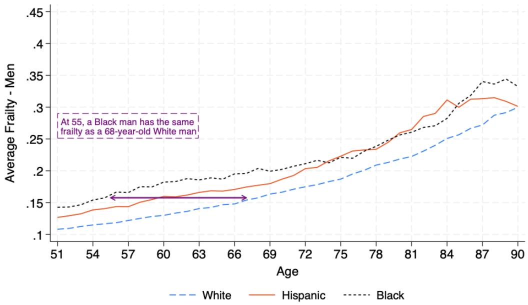
Table 1: Health deficits

| Deficit | Deficit |
|---|---|
| ***ADLs*** | Difficulty lifting a weight heavier than 10 lbs |
| Difficulty bathing | Difficulty lifting arms over the shoulders |
| Difficulty dressing | Difficulty picking up a dime |
| Difficulty eating | Difficulty pulling/pushing large objects |
| Difficulty getting in/out of bed | Difficulty sitting for two hours |
| Difficulty using the toilet | |
| Difficulty walking across a room | ***Diagnoses*** |
| Difficulty walking one block | Diagnosed with high blood pressure |
| Difficulty walking several blocks | Diagnosed with diabetes |
| | Diagnosed with cancer |
| ***IADLs*** | Diagnosed with lung disease |
| Difficulty grocery shopping | Diagnosed with a heart condition |
| Difficulty making phone calls | Diagnosed with a stroke |
| Difficulty managing money | Diagnosed with psychological or psychiatric problems |
| Difficulty preparing a hot meal | Diagnosed with arthritis |
| Difficulty taking medication | |
| Difficulty using a map | ***Healthcare Utilization*** |
| | Has stayed in the hospital in the previous two years |
| ***Other Functional Limitations*** | Has stayed in a nursing home in the previous two years |
| Difficulty climbing one flight of stairs | |
| Difficulty climbing several flights of stairs | ***Addictive Diseases*** |
| Difficulty getting up from a chair | Has BMI larger than 30 |
| Difficulty kneeling or crouching | Has ever smoked cigarettes |

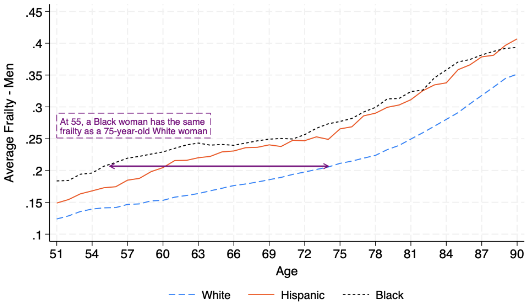
Notes: Each deficit takes a value of 0 (if the respondent reports not having it) or 1 (if the respondent reports having it).

Source: Russo et al., NBER working paper

# Average frailty rises exponentially with age



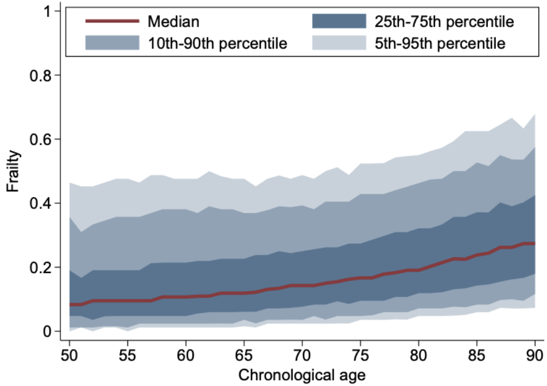

(a) Average frailty. Men



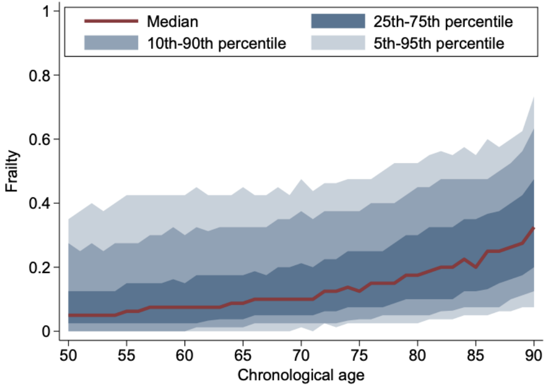

(b) Average frailty. Women

Source: Russo et al., NBER working paper

# Frailty varies more *within* age than *across* it



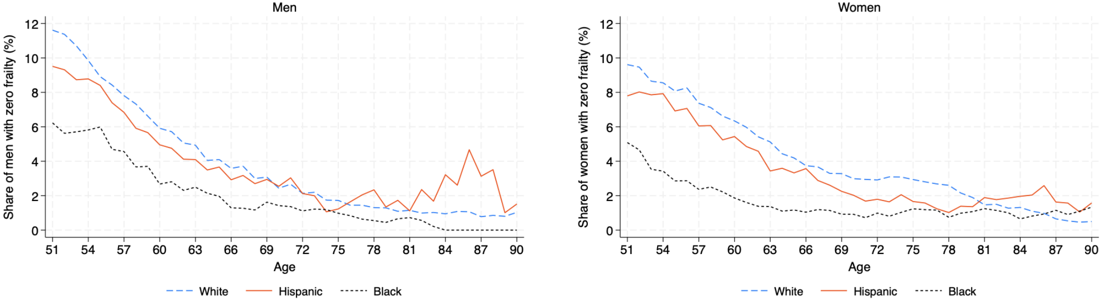

(a) Frailty distribution by age: United States




(b) Frailty distribution by age: England

Source: Kotschy, Bloom, Scott, Annual Review of Economics (2025)

# Large differences in frailty index across ethnicity




(c) Share with zero frailty. Men

(d) Share with zero frailty. Women

Source: Russo et al., NBER working paper

Table 2: Pseudo-$R^2$ table

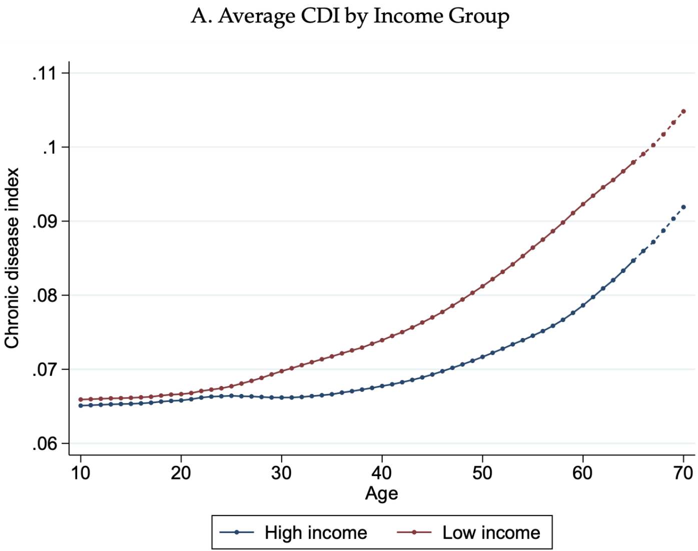
| | | Women | | | Men | | |
|---|---|---|---|---|---|---|---|
| | | White | Hispanic | Black | White | Hispanic | Black |
| SDI Recipient Next Wave | Basic Controls | 0.048 | 0.046 | 0.036 | 0.045 | 0.022 | 0.032 |
| | SRHS | 0.212 | 0.122 | 0.129 | 0.186 | 0.112 | 0.122 |
| | Frailty | 0.244 | 0.193 | 0.185 | 0.245 | 0.222 | 0.175 |
| | Frailty and SRHS | 0.268 | 0.202 | 0.199 | 0.264 | 0.241 | 0.196 |
| SS Benefits Recipient Next Wave | Basic Controls | 0.118 | 0.081 | 0.083 | 0.134 | 0.101 | 0.120 |
| | SRHS | 0.128 | 0.110 | 0.102 | 0.140 | 0.128 | 0.126 |
| | Frailty | 0.126 | 0.091 | 0.097 | 0.142 | 0.112 | 0.139 |
| | Frailty and SRHS | 0.132 | 0.123 | 0.114 | 0.147 | 0.145 | 0.145 |
| NH Entry Next Wave | Basic Controls | 0.241 | 0.172 | 0.169 | 0.220 | 0.144 | 0.122 |
| | SRHS | 0.285 | 0.209 | 0.206 | 0.266 | 0.194 | 0.176 |
| | Frailty | 0.315 | 0.231 | 0.214 | 0.303 | 0.272 | 0.234 |
| | Frailty and SRHS | 0.319 | 0.250 | 0.227 | 0.308 | 0.291 | 0.244 |
| Death Next Wave | Basic Controls | 0.166 | 0.157 | 0.120 | 0.140 | 0.157 | 0.109 |
| | SRHS | 0.240 | 0.194 | 0.169 | 0.219 | 0.212 | 0.151 |
| | Frailty | 0.266 | 0.221 | 0.189 | 0.237 | 0.244 | 0.176 |
| | Frailty and SRHS | 0.276 | 0.230 | 0.201 | 0.251 | 0.253 | 0.182 |
| | | *Percentage change from basic controls* | | | | | |
| SDI Recipient Next Wave | SRHS | 341% | 166% | 260% | 318% | 412% | 283% |
| | Frailty | 407% | 320% | 416% | 450% | 916% | 449% |
| | Frailty and SRHS | 458% | 341% | 454% | 492% | 1,005% | 514% |
| | | *Percentage change from basic controls* | | | | | |
| SS Benefits Recipient Next Wave | SRHS | 9% | 37% | 23% | 5% | 27% | 5% |
| | Frailty | 7% | 13% | 17% | 6% | 11% | 16% |
| | Frailty and SRHS | 12% | 53% | 38% | 10% | 43% | 21% |
| | | *Percentage change from basic controls* | | | | | |
| NH Entry Next Wave | SRHS | 18% | 21% | 22% | 21% | 35% | 44% |
| | Frailty | 31% | 34% | 27% | 38% | 89% | 92% |
| | Frailty and SRHS | 32% | 45% | 34% | 40% | 102% | 102% |
| | | *Percentage change from basic controls* | | | | | |
| Death Next Wave | SRHS | 45% | 24% | 41% | 57% | 35% | 39% |
| | Frailty | 60% | 41% | 57% | 69% | 55% | 62% |
| | Frailty and SRHS | 66% | 47% | 67% | 79% | 61% | 61% |

Source: Russo et al., NBER working paper

# The Chronic Disease Index (Danesh et al., 2025)

- **Data:**
  - *Context:* Admin data from the Netherlands (17m pop, 20 years).
  - *Inputs:* Pharmacy records (objective) mapped to 22 chronic conditions (diabetes, mental health, CV, etc.).
  - *Aggregation:* Use ML to weight conditions based on their predictive power for **mortality at age 70**.
- **The Chronic Disease Index (CDI):**
  - Captures the "mortality-weighted" burden of disease at any age.
  - Can answer: at what age does a high-income person have the same CDI as a low-income person?
- **Key findings:** health inequality opens up early.
  - 50% of the income-health gap at age 70 is already visible by age 40.
  - Rate of chronic disease accumulation differs a lot by income; big driver of gap at 70 (vs. other forces like sorting)

# CDI gap present at age 70 already opens up at age 40



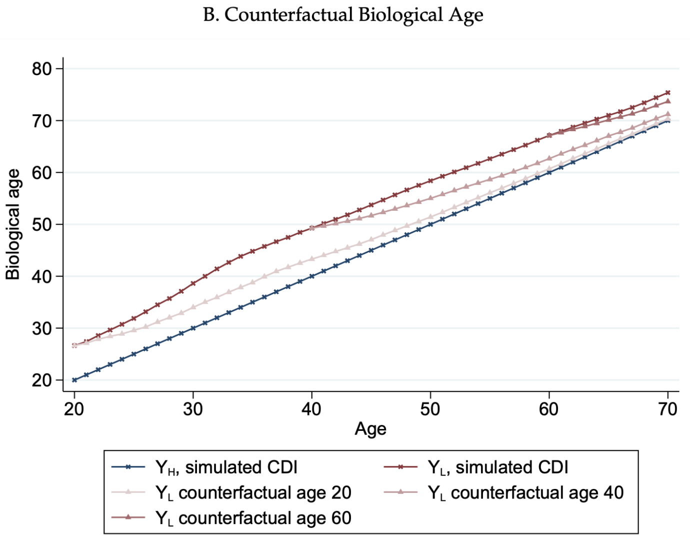

Source: Danesh et al., NBER working paper

- For group $Y$ (e.g. low or high income), define:

$$\Delta CDI_{Y,a} = E(CDI_{i,a+1} \mid Y_{i,a+1} = Y) - E(CDI_{i,a} \mid Y_{i,a} = Y)$$

- Decompose into:
  - **Biological ageing:** within-individual increase in CDI conditional on staying in the same group.
  - **Sorting:** changes in group composition due to health affecting income (and vice versa).
- **Result:** faster ageing (higher incidence of new chronic conditions) for low-income groups explains a large share of the CDI gap.

# Intervention at age 40 sufficient to almost entirely close gap at 70




Source: Danesh et al., NBER working paper

# Chronological vs. biological age

- Chronological age: years since birth.
- Biological age: inferred from biomarkers, physiology, or molecular data.
- At the same chronological age:
  - Some individuals look and function "younger".
  - Others look and function "older".
- **Relevance for us:** can biological age better explain behaviour and outcomes than chronological age?

# Epigenetic clocks

- Measure DNA methylation at many CpG sites (specific sites on our DNA).
- Predict chronological age:

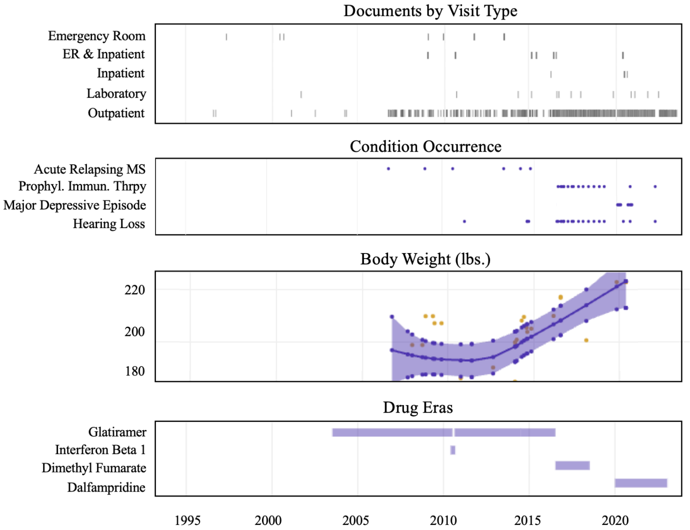
$$\widehat{Age}_i^{\text{DNAm}} = f(\text{methylation}_i)$$

- Famous clocks:
  - Horvath, Hannum, GrimAge, DunedinPACE.
- Strong predictors of mortality, disease, and functional decline.

## Alternative data sources & modern ML / LLMs in health measurement

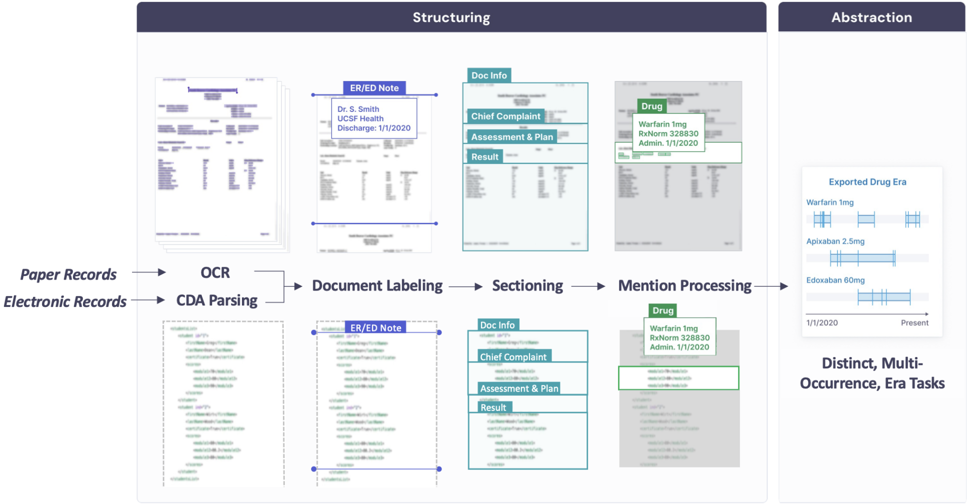
- **Consumer devices:**
  - Steps, heart rate, heart rate variability (HRV).
  - Sleep duration and staging.
- **Behavioural traces:**
  - Speech, typing, and interaction patterns as early markers of cognitive decline or mental health changes.
- **Modern ML**: deep learning on ECG, imaging, wearable streams.
  - Predict silent disease, frailty, and mortality risk.
- **Text**: large language models (LLMs)
  - A lot of medical information is captured in **unstructured** medical records/notes
  - LLMs can help extract diagnoses, trajectories, and risk factors from such notes.

# Extracting information trapped in unstructured medical notes




Source: Porter et al., "LLMD: A Large Language Model for Interpreting Longitudinal Medical Records", preprint

# Lots of data science challenges to overcome




Source: Porter et al., "LLMD: A Large Language Model for Interpreting Longitudinal Medical Records", preprint

# From measurement to welfare

- We have multiple candidate measures:
  - SRH, biomarkers, frailty indices, CDI, QALYs

- Welfare questions:
  - What do we want to maximise? Life expectancy, healthy life expectancy, QALYs, something else?

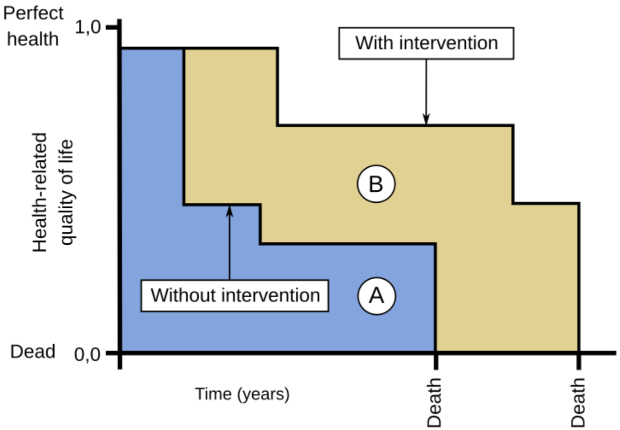
- Measurement choices influence welfare analysis, policy evaluation, and design of healthcare systems.

- **Real-world example:** National Institute for Health and Care Excellence (NICE) use approx .£20-30K per QALY threshold to approve medicines/devices/procedures for NHS.

## QALYs and DALYs

- **QALY:** quality-adjusted life year.
  - 1 year in perfect health $=$ 1 QALY.
  - 1 year at quality weight $q = 0.6 = 0.6$ QALYs.
- **DALY:** disability-adjusted life year.
  - Years of life lost (YLL) $+$ years lived with disability (YLD).
  - Measures *burden* rather than welfare.
- QALYs/DALYs used to:
  - Health technology assessments (HTAs)
  - Measuring health sector productivity
  - Analyzing disease burdens

# Quality-adjusted life year




By Jmarchn — Own work, CC BY-SA 3.0, Wikimedia Commons

**Question:** What important considerations might be missing from QALY measures, when a) making health budget allocation decisions, or b) assessing cost-effectiveness of treatments?

# NICE health technology assessment (HTA) principles

**"NICE does not set the budget for the NHS. The objective of NICE's evaluations is to offer guidance that represents an efficient use of available NHS...resources. For these reasons, the reference-case perspective on costs is that of the NHS...Productivity costs should not be included."**

# A simple (somewhat contrived) model for health technology assessments

**Two technologies**

- **Old-tech** $(O)$: improves health of old by $\Delta H_O$; cost $k_O$ per patient; no effect on productivity.
- **Young-tech** $(Y)$: improves health of young $\Delta H_Y < \Delta H_O$ cost $k_Y > k_O$ per patient; raises productivity (and hence earnings and tax revenue).

**NICE perspective**

- Evaluates technologies using health gains $\Delta H$ and NHS costs $k$ only.
- Indifferent to distributional concerns: young and old valued equally.
- Productivity gains are excluded from the calculation.

# NICE's decision rule

**Incremental cost-effectiveness ratio (ICER)**

$$\mathrm{ICER}_j = \frac{k_j}{\Delta H_j}, \quad j \in \{O, Y\},$$

with $\Delta H_O = \Delta H_Y = \Delta H$.

**NICE ranking**

$$\mathrm{ICER}_O = \frac{k_O}{\Delta H} < \mathrm{ICER}_Y = \frac{k_Y}{\Delta H} \quad \Rightarrow \quad \text{Old-tech preferred, Young-tech rejected.}$$

- The productivity benefit of Young-tech is invisible to the decision rule.

### Remaining NHS budget

- Let $B$ be the remaining NHS budget available for new technologies.
- Two candidate technologies: Old-tech and Young-tech as before.

### Budget competition

- Suppose the budget is tight enough that:

$$k_O \leq B < k_O + k_Y.$$

- If budget is fixed, only one technology can be funded.
- Assume Young-tech generates additional tax revenue $\Delta T$ by raising productivity:

$$\Delta T = \text{extra NHS resources from higher earnings.}$$

### Consider a world where:

$$k_O + k_Y \leq B + \Delta T.$$

Young-tech's productivity-induced fiscal feedback enlarges effective health resources enough to fund both technologies.

**NICE decision**

- With $\Delta H_O > \Delta H_Y$ and $k_O < k_Y$, Old-tech option is preferred.
- In the baseline case:

Adopt Old-tech only, reject Young-tech.

- Societal gains from increase productivity from Young-tech are left on the table.

# Extended evaluation rule

**Extended decision**

- Suppose we instead adopt Young-tech first.
- Immediate cost: $k_Y \leq B$.
- Young-tech raises productivity, generating extra tax revenue $\Delta T$.

**Fiscal expansion condition**

$$k_O + k_Y \leq B + \Delta T.$$

- After adopting Young-tech, the NHS now has sufficient resources to also fund Old-tech.

**Pareto comparison**

- Old: same health as in the NICE baseline.
- Young: better health *and* higher productivity/earnings.
- No one is worse off; some are strictly better off gives **Pareto improvement** relative to the NICE choice of Old-tech only.

# Retirement, disability, and biological age

- Many policies hinge on chronological age cut-offs:
  - Statutory retirement age.
  - Eligibility for disability, pensions, long-term care.
- Frailty and CDI evidence:
  - Large heterogeneity in health at a given age.
  - SES gradients in ageing speed.
- **Open question:** should eligibility depend (in part) on health/biological age?

# Key takeaways from today

- Health and ageing are **multidimensional** and **latent**.
- Grossman and Strulik/Dalgaard offer different core theories of health dynamics.
- Measurement is hard:
  - SRH: subject to biases
  - Objective measures: need to take stance of mapping from $N$-dimensions measure to low-dimensional measure of health
- **New data and AI** are creating opportunities
- Careful measurement is a prerequisite for credible empirical work and sensible policy on health and ageing